Chapter I

INTRODUCTION

1.1 MEANING OF LIFE TIME AND SCOPE OF SURVIVAL ANALYSIS

Life Time

In biological and medical studies an important variable of interest is the duration of time for which an individual survives. In engineering, the total duration for which the system works satisfactorily is an important characteristic. In general a living object ultimately experiences death and a working unit ultimately experiences the failure. Such non-negative variables play an important role and in the literature these are studied in depth. A non-negative random variable is referred to as "life time", "survival time" or "time to failure".

The following are some related terms used to define life time

1) Time origin (Starting time).

2) Measurement scale for the time.

3) Ending event.

Consider the following examples to describe the above terms.

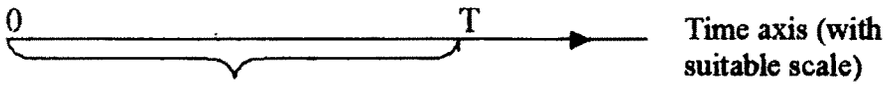
1. To study the life times of human beings or animals, origin is epoch of birth, measurement scale for time is days, months or years and ending event of interest is death of an individual.

To study the behavior of time to recover from a disease, epoch of start of treatment or date of admission in a clinical trial is origin, measurement scale is days, months or years and ending event, is relief from the disease or symptoms. Note that, in this situation life time denotes time duration between the beginning of the treatment and relief from the disease.

2. In industry, to study the behavior of the life times of certain components, the components have to be tested under the environment, for which they are designed. Here the event of interest is the non-functioning of a component and associated life time or the survival time is the period for which the component works.

3. In the study of unemployment, one might be interested in the time for getting an appointment or job. In this case, the time at which a person starts for searching a job is the origin and getting a job is an ending event. Further the time between the epoch of beginning of search and that of getting a job is the period of interest.

Thus diagramatically life time can be represented as



Starting time life time Ending event.

Survival analysis

Survival analysis is an important branch of Statistics, which deals with the analysis of data on life time. The data may be from any field like medical research, economics, demography, reliability, biological and physical science etc.

For the analysis of the data, a suitable statistical model has to be considered. Using standard classes of parametric distributions we can develop such model. The behavior of life time might depend on the concomitant variables and to develop models in such situations the concept of regression can be used. Based on the hazard function (failure rate function) some popular models have been developed, namely proportional hazards model, accelerated hazards model. Also by using survival function an accelerated failure time model is developed. However in practice, data obtained may not be complete in the sense that it might be a censored data. Techniques have been developed to analyse the data in such situations.

In the following, we describe in brief, regression models for survival analysis in Section 1.2.

1.2 REGRESSION MODELS FOR SURVIVAL ANALYSIS

In this Section, we discuss how one can develop life time models using the concept of regression. Regression analysis is a statistical technique for investigating and modelling the relationship between variables. Applications of regression are numerous and occur in almost every field, including engineering, physical science, economics, management, life and biological sciences.

A simple model to study dependence of one variable say T (also called response variable) on other variable (or concomitant variable) say Z is a linear regression model given by,

$$T = \beta_0 + \beta_1 Z + e, \tag{1.2.1}$$

where β_0 and β_1 are the parameters of the model and e is an error component. Generally least square methods are used for fitting of a model.

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the estimates of β_0 and β_1. Let $\hat{T} = \hat{\beta}_0 + \hat{\beta}_1 z$ be the estimated value of T corresponding to the value z of the concomitant variable Z. It is to be noted that the difference between observed and estimated value, that is $T - \hat{T}$ is an error, which might be due to natural fluctuations.

Let the response variable T be related to k concomitant variables $Z_1, Z_2, \ldots, Z_k$ by the relation,

$$T = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_k z_k + e. \tag{1.2.2}$$

As there are more than one concomitant variables and RHS is a linear function of $z_1, z_2, \ldots, z_k$. This model is known as multiple regression model.

One simple way to study the dependence of life times on concomitant variables is through above regression models, in which life time (T) has a distribution that depends upon the concomitant variables. In survival analysis, generally variable of interest is lifelength of the individuals and concomitant variables are special characteristics of the individuals such as age, sex, disease, dose of medicine etc.

To study the heterogeneity present in the population one can divide the population into strata based on various levels of a suitable concomitant variable. In clinical trials to study life time of a patient concomitant variables may be taken as treatments, or intrinsic properties of the individuals etc. The terms treatments, intrinsic properties are described below.

1. **Treatments:** In biological studies, treatment is one of the concomitant variables. To compare two treatments say a new treatment with the "control" or standard treatment we define a binary concomitant variable Z as follows,

$$Z = \begin{cases} 1, & \text{if an individual receives the 'new' treatment.} \\ 0, & \text{if an individual receives the 'control' treatment.} \end{cases}$$

To compare k treatments one can define concomitant variable Z taking values 0,1, 2, ..., k-1 or may be 1, 2, ...k.

2. **Intrinsic properties:** Intrinsic properties describe a medical history prior to the admission to the study. For example age, general physical conditions, the type of disease (like blood pressure, diabetes) of the patient or it may be the duration for which a patient is suffering from a disease.

In this dissertation, we study the life time models where the life time T depends on the concomitant variable. We consider a proportional hazards model which is a distribution free regression model in the sense that $h_0(t)$ (baseline hazard function) is arbitrary. This model was introduced and developed by Cox (1972). The Cox model is very widely used in data analysis especially in medical research. A good historical discussion on this model is given in Miller (1981), Lawless (1982), Cox and Oakes (1984), Collect (1994), Smith (2002). Another model in which concomitant variables are used to accelerate (or to retardate), the wearing out process popularly known as accelerated failure time model is discussed. This model is described via a baseline survival function. Further, Chen and Wag (2000) have proposed an accelerated hazards model by using baseline hazard function and transformation of variable involving a parameter. Some other failure time models such as additive model, model based on minimum of life times, transferred origin models have been defined.

In the following Section we shall introduce the concept of censoring as well as different types of censoring.

1.3 CONCEPT OF CENSORING

Censoring is a technique used to reduce the time of experiment and / or the cost of the experiment. In this method, the sampled observation may contain only a partial information about the random variable of interest. Such incomplete observations on the failure time are known as censored observations. A sample under such a set up is referred as a censored sample. In survival analysis, the values assumed by a life time variable are usually quite large. In such cases the experiments have to be modified so that time and cost restrictions are satisfied. While doing so, some information may get lost.

Following are the types of censoring.

1.3.1 Type I Censoring (Time Censoring)

Consider a life testing experiment in which n identical items are placed on test and the study is discontinued at a predetermined time t_0 (say). Exact life times are known for those items failed prior to t_0. The exact life times of those items survived by time t_o are not known, however for these items it is known that these life times are greater than t_0. Let n_u be the number of items failed by time t_0. These items are termed as uncensored items. The remaining items n_c =

$n - n_u$ items remain operative at time t_0 and they are known as censored items (n_c). In this case the data consists of the life times of the n_u failed items and the censoring time t_0 for the n_c censored items.

1.3.2 Type II Censoring (Order Censoring)

In this case n identical items are simultaneously put into operation. The study is discontinued when a predetermined number k (< n) items fail. Here the data consists of the life times of the k items that failed (at the time epochs say $(x_{(1)}, x_{(2)}, x_{(3)}, \ldots x_{(k)})$) and the fact that (n-k) items are survived beyond $x_{(k)}$.

1.3.3 Random Censoring

There are two types of random censoring techniques discussed as follows.

i) Right Random Censoring : It occurs when the complete life times are not observed for many reasons, which are beyond the control of the experimenter. It occurs in any one of the following situations:

1. Lost to follow up : A patient may decide to move elsewhere then the experimenter may not be able to observe the patient again. The only information available on the survival experience of that patient is the last date on which patient was known to be alive. This date will be the last time that the patient reported to a clinic for a regular check-up.

2. **Withdrawal (leave) from the study :** The patient may not afford the expenses, treatments or medicines are ineffective or treatments produce bad side effects so it may become necessary to discontinue the treatment.

3. **Termination of the study :** The patient may not be satisfied with the effect of the given treatment.

ii) **Left Censoring :** Left censoring occurs, if the actual survival time of an individual is less than that observed. It may or may not be identical for all observations. It occurs less frequently than right censoring.

The data set may contain both left and right censored observations. Following is an example from Miller (1981). A psychiatrist travelled to Africa to determine the age at which children have learnt to perform a particular art. The life time is the age at which the child learn to perform a particular art. Those children who already knew how to perform an art were left censored. Further, those who did not learn the art even by the time he leaves the village, were right censored observations and those who learn the art during the period of stay of the psychiatrist are uncensored observations.

Remark (1.1) : In type I censoring the time of termination is fixed while the number of failures before time t_0 is a random variable whereas in type II censoring case the situation is reverse in that, k the number of items that failed is fixed while $x_{(k)}$ the time at which the experiment is terminated is a random variable.

While in random censoring, the numbers of censored as well as uncensored observations are random and time for which the study ends may also be random.

The Chapter I is introductory, describes meaning of life time and scope of survival analysis. Also some regression models considered in the dissertation and the concept of censoring are described in this Chapter.

We now present the Chapterwise summary of the dissertation.

1.4 CHAPTERWISE SUMMARY

Starting with some basic concepts related to life time distributions necessary definitions, interrelationships between the density function, survival function, cumulative hazard function are given. Section 2.3 covers relationship between the survival function and hazard function for the discrete distributions. Such a relationship is established for the mixture of continuous and discrete distributions in the last Section. Some examples are given to illustrate the result.

Chapter III, deals with the proportional hazards model. A test of significance for the regression parameter (β) based on the Rao's score test statistic is described based on the conditional likelihood. Further, the marginal likelihoods for uncensored and censored data are discussed. In Section 3.5 a test for β in the presence of ties based on the likelihoods proposed by Cox (1972) and approximate likelihood suggested by Kalbfleisch and Prentice

(1972) is given. Next Section is devoted to two sample problem for testing equality of two survival functions under ties. At the end of this Chapter, the solution for the two sample problem is applied to a real data set on two types of treatments on gastric cancer.

Chapter IV is devoted to accelerated life time models. We introduce an accelerated failure time model. Some accelerated failure time models are described by using log-logistic, Weibull distributions. We describe an accelerated failure time model with time dependent concomitant variables. This Chapter also covers accelerated hazards model and estimation procedure of the regression parameter β. Finally, the Chapter ends with discussion on some other failure time models. The list of references is given at the end.